DOCUMENT RESUME

ED 095 815 95 IR 000 918

TITLE	Proposed Research Program (Part A: Technical Narrative) on Assessment and Analysis of Education Satellite Communications Demonstration. Phase III.
INSTITUTION	Battelle Memorial Inst., Columbus, Ohio. Columbus Labs.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	RFP-NE-R-74-0002
PUB DATE	21 May 74
CONTRACT	NIE-C-74-0045
NOTE	67p.
EDRS PRICE	MF-$0.75 HC-$3.15 PLUS POSTAGE
DESCRIPTORS	Communications; *Communication Satellites; Evaluation Methods; *Program Evaluation; *Research Design; Research Methodology; Research Problems; Research Proposals; Scientific Principles; Statistical Analysis; *Telecommunication
IDENTIFIERS	ATS F Satellite

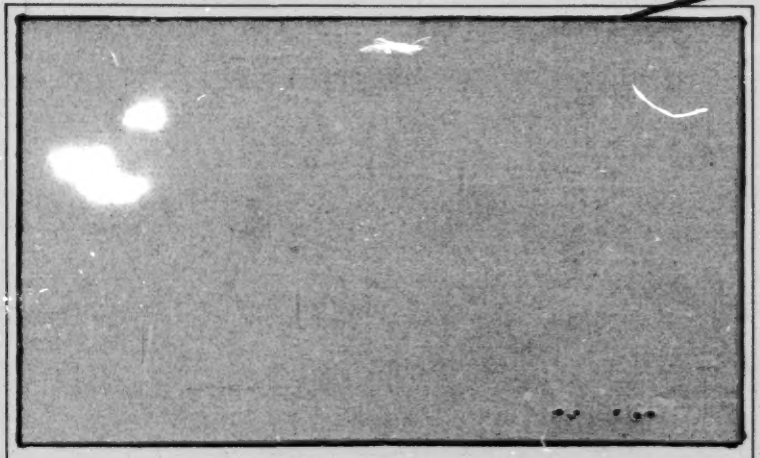
ABSTRACT

A proposal to monitor and summarize the experience that will be gained during the course of the ATS-F Education Satellite Communications Demonstration is described. The goals of the demonstration and the context in which it must be evaluated are discussed in an introduction. Subsequent sections deal with specific analytic approaches proposed, the project design, and the organization of the plan. The section on project design includes discussions of both the report schedule and the category system adopted for the analysis. The project management plan and the work plan and schedule are detailed in the organizational section. Finally, there are a set of four appendixes, entitled: 1) design of large-scale experiments; 2) data collection and analysis; 3) potential concept papers; and 4) target populations and the "assessment" process. The first appendix is useful in the planning of any large scale experiment, since it describes statistical concepts in general terms, and is not restricted to ideas relevant only to this study. (WDR)

ED 095815

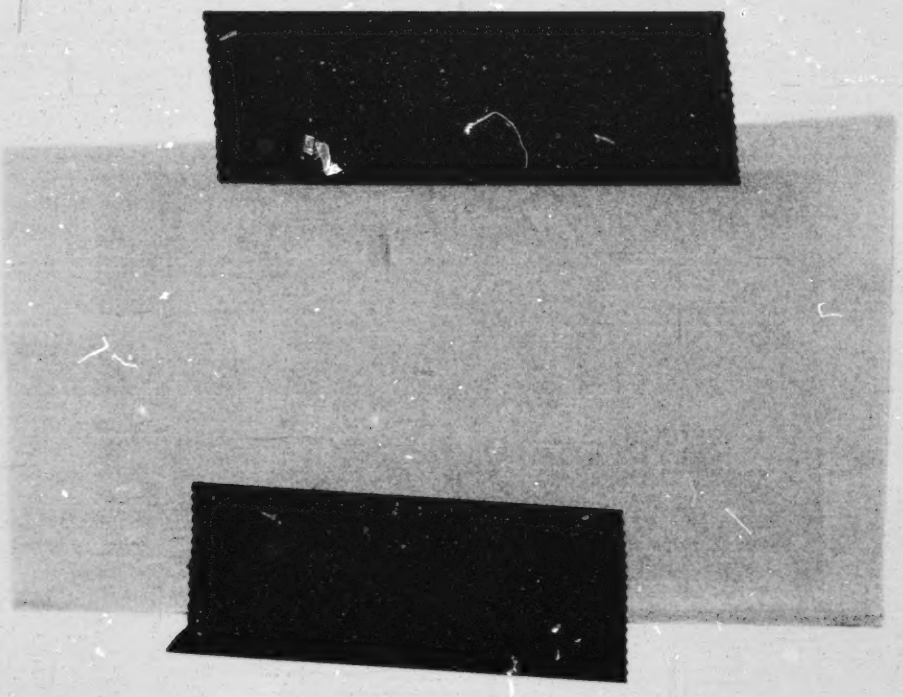
PART A

PROPOSED RESEARCH PROGRAM



Battelle
Columbus Laboratories

IR 000 918



BATTELLE'S COLUMBUS LABORATORIES comprises the original research center of an international organization devoted to research and development.

Battelle is frequently described as a "bridge" between science and industry — a role it has performed in more than 90 countries. It conducts research encompassing virtually all facets of science and its application. It also undertakes programs in fundamental research and education.

Battelle-Columbus — with its staff of 2500 — serves industry and government through contract research. It pursues:

- research embracing the physical and life sciences, engineering, and selected social sciences
- design and development of materials, products, processes, and systems
- information analysis, socioeconomic and technical economic studies, and management planning research.

505 KING AVENUE • COLUMBUS, OHIO 43201

ED 095815

PROPOSED RESEARCH PROGRAM
(Part A: Technical Narrative)

on

ASSESSMENT AND ANALYSIS OF EDUCATION SATELLITE COMMUNICATIONS DEMONSTRATION PHASE III

CONTRACT NO. NIE-C-74-0045
(RFP NO. NE-R-74-0002)

to

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
NATIONAL INSTITUTE OF EDUCATION

May 21, 1974

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BATTELLE
Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

PREFACE

This proposal consists of three separately bound volumes:

PART A - TECHNICAL NARRATIVE

PART B - STATEMENT OF QUALIFICATIONS

PART C - BUSINESS PROPOSAL.

This volume is Part A - Technical Narrative.

TABLE OF CONTENTS

	Page
PART A: TECHNICAL NARRATIVE	1
INTRODUCTION	1
AN ANALYTIC APPROACH	6
THE PROJECT DESIGN	9
Report 1	10
Men	11
Machines	17
Materials	18
Money	19
Report 2	21
Report 3	26
Summary Final Report	27
Supplementary Publications	27
ORGANIZATIONAL APPROACH	32
Project Management Plan	32
Work Plan and Schedule	37

APPENDIX A

DESIGN OF LARGE-SCALE SOCIAL EXPERIMENTS A-1

APPENDIX B

DATA COLLECTION AND ANALYSIS B-1

APPENDIX C

POTENTIAL CONCEPT PAPERS C-1

APPENDIX D

TARGET POPULATIONS AND THE "ASSESSMENT" PROCESS D-1

PROPOSED RESEARCH PROGRAM
(Part A: Technical Narrative)

on

ASSESSMENT AND ANALYSIS OF
EDUCATION SATELLITE COMMUNICATIONS DEMONSTRATION
PHASE III

to

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
NATIONAL INSTITUTE OF EDUCATION

from

BATTELLE
Columbus Laboratories

May 21, 1974

PART A: TECHNICAL NARRATIVE

INTRODUCTION

This study must provide information for the future planner of educational systems.

In the following pages we outline a plan to monitor and summarize the experience that will be gained during the course of the ATS-F Education Satellite Communications Demonstration. The plan is intended to both collect the most useful experience and to place it in a context where it will be most helpful to the planners, educators, and public who may consider further experiments and applications.

Most importantly, the plan reflects a firm understanding of the roles of participants and observers, and the nature of scientific method as applied to large-scale social experiments. The forthcoming satellite exercise cannot be considered a "test"; neither

is it a "demonstration"; nor, in the conventional sense, an "experiment". Rather, it is an exploration. Three separate regional organizations, each with its own problems, motives, and goals, are exploring the application of a new and untested delivery system. The three regions have little in common besides the satellite, relatively remote, dispersed populations, and an interest in education. With neither common needs nor prior experience, and a purely exploratory intent, their goals and plans have been more adaptive and expedient than deliberate, and they will probably continue as such.

As this unique educational delivery system is explored in three areas . . .

This is as it should be. And words like "evaluation" and "assessment" are neither desirable nor appropriate for this exploration. Rather, the situation calls for the presence of a friendly, knowledgeable, and neutral observer who can recognize and record both what happens and why . . . relating this record to general problems of educational delivery so that it may be considered and applied in the future by educational systems designers.

An unbiased overview is needed for future planners.

There is, of course, a sharp contrast between this need for overview and the interests and goals of the three regional assessment teams. As participants, they are deeply involved in questions of the quality and educational impact of their material. They are intimately and appropriately involved with the target population and have an urgent and continuing need for feedback that can help to "tune" the educational content. Their viewpoint as participants carries the weight and authority of direct contact. But it cannot, by definition, ever be neutral and detached.

Each regional team is a concerned participant, neither neutral nor detached.

Similarly, in this exploratory environment, the National Institute of Education (NIE) is hardly a non-participant. Throughout the course of the satellite plans, the Federal Government has played an active role . . . funding, mediating, advising, encouraging, and criticizing. This is an appropriate role; for, in our opinion, it is a principal function of NIE to explore and describe alternative educational methods, tools, and delivery systems so that they may be considered by the public and the political and educational communities in the light of their own goals.

Likewise, NIE is a deeply involved participant.

To perform this function effectively, it is both appropriate and necessary for NIE to establish a neutral observer with sophistication in the mechanics and problems of televised education but with no role whatsoever in the project outcome. We propose this as the role of Battelle-Columbus in the Education Satellite Communications Demonstration (ESCD).

BCL's role is purely that of an observer ...

To maintain this perspective and neutrality, we must remain observers and resist the temptation to "get in the game", whether in the guise of "measurement", "advice", or of any other excuse for direct involvement with the target population.

... analyzing but not advising, extrapolating but not participating.

* * * * * * * *

Far more fundamental to this effort than a clear definition of roles, however, is the nature of scientific method as applied to a large-scale social experiment. We suggest that a misguided interpretation of scientific method has often led to an inability to draw valid conclusions from social experiments. Where

a target population can be carefully structured and isolated from extraneous influences, the word "experiment" has classic significance, but the results are limited.

Applying scientific method to such large-scale social experiments is difficult and conventional techniques inapplicable.

In large-scale social experiments the variables are so numerous . . . the populations so diverse . . . that procedures based upon "statistically valid" samples are not only inapplicable but also lead inevitably to confusing and negative conclusions. Such attempts lead to "We did not observe any statistically significant results."

One could not increase the statistical precision of the Rocky Mountain sample by adding the students of Alaska and Appalachia. Each new group of students simply adds a new and imponderable set of variables, actually decreasing the precision. While this is an extreme example, the same problem occurs when we attempt to add the students of additional Indian villages or any such superficially similar groups. What then is the role of scientific method in such large-scale exploration?

The essence of scientific method must be understood and applied.

The essence of scientific method is the ability to make valid predictions. This ability is both the goal and the test of the technique employed . . . its ultimate sophistication is found in the detailed and quantifiable predictions of the physical sciences, where sample sizes are large and individual variations small. But in social experiments such as the satellite exploration, where the populations are small and the variations large, the ability to make valid predictions is no less a proper criterion of scientific validity.

* * * * * * * *

Thus, in this study there will be no control groups . . . there will be no "evaluation" . . . and there will be no attempt to assess the "educational impact".

Rather, there will be an exceptionally serious and scientifically validated interpretation of its potential, as well as the strategy applied in its utilization. There will be an extensive comparative analysis of the satellite delivery system as demonstrated in three differing regions, presenting different populations, different goals, and different application strategies. And there will be a publication of the real-world experiences and lessons as interpreted by the key participants themselves.

The output of this study will be a synthesis of the experience of three significantly different explorations.

The satellite is one alternative educational delivery system. It will be considered and weighed by educational planners, politicians, teachers, media planners, and the public for many years. It is our goal to supply the most valid, complete summary for their appraisal.

AN ANALYTIC APPROACH

Scientific method demands that we
Measure
Relate
Predict
Verify.

In previous concept papers and lectures we have summarized scientific method as *Mister PV*, shorthand for Measure, Relate, Predict, and Verify. (For details see Appendix A.) Applied to the satellite experiment this method will encompass a variety of descriptive techniques--quantitative where possible, but qualitative or hierarchical where necessary, with data gathered systematically, related and explained to the best of our ability, and constantly tested to extrapolate the course of events.

Our analysis will interrelate
Men
Machines
Materials
Money ...

For convenience of description and analysis, we have categorized the delivery system in four major, interrelated components--Men, Machines, Materials, and Money. While these are intricately related and cannot be truly isolated, we believe that this categorization will provide a convenient and useful device for the educational system planner who must consider the satellite and other alternative delivery systems. Together, these four categories and their relationships provide a description of the system at any point in time. By a process of periodic updating of each area (as well as of their relationships), we can document the detailed course of the experiment.

... in fact, and as <u>predictions</u> for the future.

We propose to do this updating at frequent intervals (no greater than one month apart). At the same time, we will attempt to explain changes that take place, and predict the future course of events. As indicated previously, our ability to make such predictions successfully is the true test of scientific validity, and the detail of such predictions indicates the depth of understanding.

We will generate two series of monthly reports:

Each month two reports will be prepared summarizing new information about the project areas: a "descriptive" report for review by NIE and the three projects, and a confidential "interpretive" report for analytical use only by the Battelle team until the final report is submitted. The descriptive report will summarize to the best of our ability (both quantitatively and qualitatively) the status of the projects in terms of the four categories. It will be especially concerned with changes, the reasons for their occurrence, and the expectations of participants. This could include, for example, changes in picture quality as the site coordinators become more familiar with the antenna or the hardware is debugged. It could report an improved attitude on the part of instructors as they become more accustomed to the mechanics, or a growing enthusiasm among the students as they become familiar with the television characters, or a change in community attitude as the *Alaska Native Magazine* becomes well known. In each case, it will be our goal to observe and note such changes . . . then, in the second report, explain their occurrence, and verify this insight through prognosis for further change.

"Descriptive", to summarize the facts.

"Predictive", to reflect and test our understanding and analysis of the facts.

To do this, each month we will prepare a second report that will record our interpretations, predictions, and a month-by-month comparison with the actual course of events. For obvious reasons, this exercise constitutes a critical and confidential test which would not be valid if it could in any way become part of the satellite experiment. Thus, in contrast to the descriptive report, which will be submitted to both the regions and the NIE staff for additions, comments, corrections, suggestions, etc., the predictive report must remain completely confidential throughout the course of the ESCD. At the termination

of the project, this month-by-month accounting will be incorporated as an appendix to the final report so that NIE and other interested observers can assess for themselves the validity of our predictions and interpretations.

This need to separate the observer's opinions from the experiment is well known. A recent review of the famous Hawthorne experiments*, for example, clearly shows this influence . . . for the simple reason that the participants were actively aware of the experimental design, data, and outcome.

Our predictions must not influence the demonstration.

And it is for this reason that we see a need for a complete isolation from involvement with the target population, and a strong need to maintain a friendly but nonparticipatory neutrality toward both the participants and NIE.

We believe that we are exceptionally well qualified to exercise this role. We have great sympathy with the fundamental goals of the satellite experiment, a sophisticated understanding of its complexity and difficulty, a broad background in related activity and policy issues, and a sincere desire to observe and explain the important aspects of the satellite experiment.

As we have said previously, there is no question of assessing the satellite's "educational value". The satellite is an alternative delivery system and it will be seriously considered in the future design of large and remote educational systems. It is our goal to provide the future designer with the most useful information about when and how it may be successfully applied.

Our goal is to build on regional results to establish a framework for the future planner.

---

* "What Happened at Hawthorne?" by H. M. Parsons. *Science*, Vol. 183, pp 922-932 (8 March 1974).

THE PROJECT DESIGN

This project plan is designed to yield a documentation and analysis that will allow a future planner or system designer to consider the use of satellite communication as one alternative educational delivery system. However, despite its size, the current Education Satellite Communications Demonstration can by no means be considered definitive. Thus, to be useful, the final report of the experiment will present not only a thorough description of what happened and why, but adequate detail so that the planner can weigh the analyses and extrapolate the experience to new and different circumstances.

The ESCD is not a definitive feasibility study . . . and many other factors will be weighed by future planners.

For this reason, we have chosen a format which relates the experiences of the Appalachian, Rocky Mountain, and Alaskan regions to a group of common denominators that are central to this and any other delivery system. During the project we will also attempt to develop a comparison of these characteristics with a number of likely alternatives.

We will relate regional experiences and consider alternatives ...

Because of this "planner's" perspective, we will restructure much of the detailed information gathered by the three regions, but we see no reason to duplicate or to become closely involved in the regional measurement programs.

But will avoid involvement in the regional programs or contact with their target populations.

Fortunately, our approach is not only appropriate, but also economical . . . and it avoids the very serious questions of impact on the target populations. Hopefully, by concentrating on basic features of the delivery tool and examining its performance when applied

to the substantially different problems and activities of the three regions, one may establish patterns and guidelines that may be extrapolated, however tentatively, to new environments.

We will do this through three principal report types. Together they will provide both the needed documentation and overview and the basis of a retrospective discussion for general publication.

The study will be recorded in three types of reports.

Report 1 - HUMAN RESOURCES, EDUCATIONAL MATERIALS, EQUIPMENT, AND FUNDING FOR EDUCATIONAL SATELLITE SYSTEMS: A Planners' Guide based on the ATS-F Education Satellite Communications Demonstration, 1974-1975

Report 1 is the information gathering and documentation portion of this project. In its monthly summaries will be statistical and narrative data on both events and opinions--the Men, Machines, Materials, and Money that make up the ESCD. In the following pages we provide a discussion of its role and content.

Report 1 documents the history, performance, and consequences . . .

This report will be prepared on a monthly basis as well as in summary form at the conclusion of the demonstration. We will also prepare an initial background summary following the same format. This initial summary will provide a review of the goals and plans of the projects stated in terms of four basic categories which are described below. Thus, this initial report will serve as a "baseline" statement of goals, while the monthly reports will be concerned with the progress and changes occurring during the experiment.

In terms of Men, Machines, Materials, and Money.

Each report will provide information in four mnemonically labelled categories chosen for their critical

interest to the educational planner and system designer: Men, Machines, Materials, and Money. For each category we will summarize the experience of each region and then attempt to combine this information so as to develop a more generalized description of the satellite potential (qualified or reinforced by the experience in the three regions).

Men

Even within the brief period of our involvement in the current planning project, it has been vividly clear that a key factor in developing and applying such a large-scale tool lies in the personality, competence, and dedication of the principal participants. Experiments of this magnitude do not launch themselves, and the subjects of each regional activity serve to emphasize that the choice of a delivery system is especially sensitive to the personalities and dynamic interactions of the participants as well as the educational and political environment. We suggest, for example, that the choice of locations is only partially due to the existence of educational need combined with low population density and difficult broadcast terrain. The locations also reflect the presence of visionary, dedicated, and determined individuals with the talent, persistence, and numbers necessary to mount such a difficult undertaking.

Men is a key cateogry, reflecting profound influence on planning and performance.

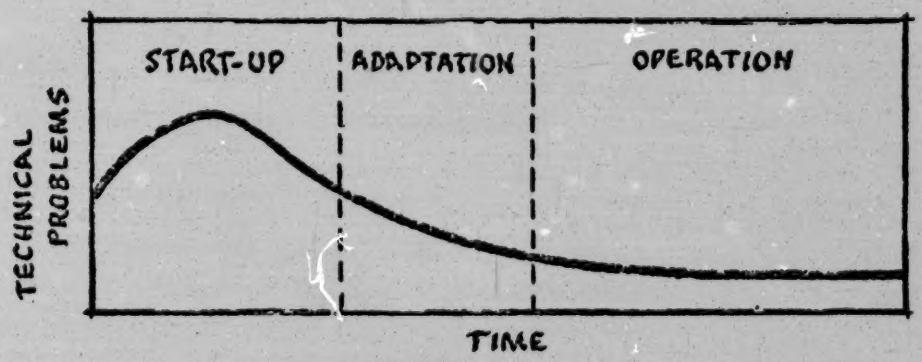
To describe the category Men consistently and systematically, we have divided it into three subgroups that reflect major functions in the demonstration:

- Planners
- Implementers
- Others.

What lessons can we learn from the planners?

Planners. These are those personalities concerned with the conception of the ATS-F demonstrations and the definition of the demonstration roles. A future planner or system designer will surely want to know about their numbers, personalities, and roles both in establishing the program and adapting to difficulties.

Recognizing that there is no simple measure of either technical difficulty or community acceptance, we hypothesize a three-phase transition in the inception of any such technical innovation.



When the device is first introduced, we can expect a rapid appearance of the inevitable technical difficulties that are fully recognized only during the course of field use. Almost immediately a process of adaptation in equipment, strategy, and perhaps goals, must begin; leading eventually to a more stable operational period.

Since such difficulties accompany any major innovation, the role of the key planning individuals and organizations will be of great interest to the future policy maker and system designer. A principal consideration in describing the regional activities will be an assessment of the status of the system in terms of these evolving roles. We can expect, for example, that as the current efforts become more stable, the planners will begin to withdraw from problem-solving and operational difficulties to concern themselves increasingly with questions of future development, system expansion, and alternative applications.

Throughout the project we will maintain a uniform descriptive format both to assure consistency, and to provide a basis of comparison among the three regions. We will identify key figures and their roles both in terms of their nature and the "level of criticality" to the project's success. We will examine the organizations that are represented and their role in the project, looking particularly for evidence of interest or withdrawal as a reflection of changing concern or changing roles. We will describe the links that exist among both individuals and organizations, either as they help or hinder the ability of individuals to solve critical problems.

What about the organizational arrangements?

For a variety of reasons, we will also look at the nature and fate of opposition that may appear--when and why it occurs; how it is handled; and how it should be weighed by a future planner. Acceptance (or cessation of active criticism) by the opponents of a new system can be a most sincere form of flattery.

... and the opposition?

To the planner the dynamics of this process may well have more significance than the immediate issues.

Implementers. In the same manner, we will methodically describe the numbers, backgrounds, and experiences of the persons directly involved in carrying out the demonstrations. Here, the numbers of individuals are larger, and the description will consider both the key individuals and the larger groups which must be described in generalized terms. As in the case of the "planners", we will record changes in the character, role, and numbers assigned to staff the projects, adapt to initial problems, and gradually reach an operational stage. As operations continue, a number of variables such as staff level, reassignment of responsibilities, and change in attitude toward conduct of the demonstrations will be noted.

The implementers reflect another side of the coin.

Clearly, a key issue in our review will be to assess when (and if) a reasonably stable operation has been achieved. It may be that during the course of a one-year demonstration, one can never expect a truly operational situation, and it may be impossible to do more than note trends, speculating on an ultimate point of stability.

Others. There are many other persons who are not directly involved in the regional programs but have a relevant interest and opinions regarding its progress. These might be the editor of a local newspaper, a state superintendent of schools, or any number of such individuals representing points of view, pressures, and interests outside the direct conduct of the projects. Through these people we may find an insight into the impact and future of the project that cannot be obtained in any other way.

... and there are those who are concerned but not direct participants.

For reasons which have already been discussed, and are detailed still further in Appendix D, it is neither appropriate nor feasible to establish direct contact with the target populations themselves. We regard ourselves as observers and analysts on behalf of future planners, and it is fully appropriate to view the target population through the eyes of those who are developing and operating the systems. We find no question regarding the dedication, honesty, and competence of the regional assessment teams, though we have every respect for the magnitude of their task.

Battelle will not work directly with the target populations of the regions.

In each case we will simply watch from a distance, requesting local assistance in obtaining data to answer a few additional questions for our purposes. As outside observers we regard the authority and judgement of the local assessment teams as paramount, and we will defer to their judgement completely in the area of data collection from the sites. At the start of the project our reporting team will visit each of the regions to work out such details, and establish the routine upon which both our data gathering and narrative reporting will be based.

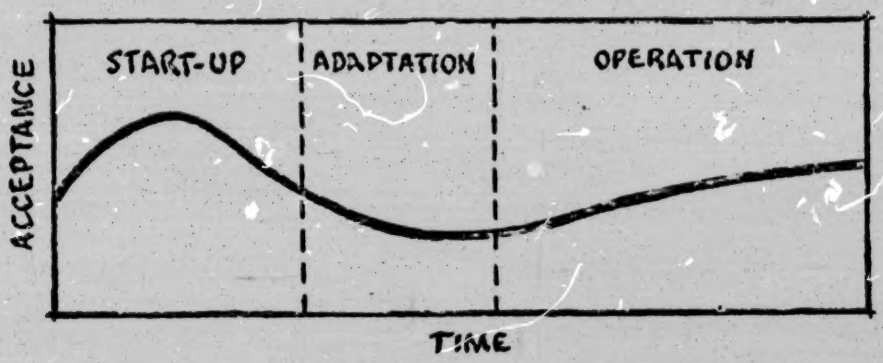
We will get target population information only via the local assessment teams.

The provision of information to Battelle by the local assessment teams represents a burden, and we will work with them not only to define the activity, but also to assess its reasonableness and cost. We propose that such costs be reimbursed through an appropriate addition to the regional funding.

The attitudes of persons concerned with the satellite and their acceptance of the program are, of course, critical to the educational planner and policy maker. Equally important, however, are the reasons

why these attitudes exist. Because of this, we will work closely with the regional assessment teams to gather direct indications of attitude and to weigh their intuitive appraisals of the reasons for their response.

We will also collect the team members' appraisals of target population attitudes.



Just as we have hypothesized a phased approach to the solution of technical problems, so we speculate on the possibility of an acceptance profile paralleling this development. Intuitively, we expect an initial enthusiasm accompanying the introduction of a glamorous new device, a great deal of public attention, and "teaching by entertainment". As problems appear and the "blush leaves the rose", one might expect an overreaction followed by a gradually stabilizing consensus as operational status is reached. One of our goals will be to either verify this view or provide an alternative model.

Generating a model of acceptance over time.

In this as in other categories of the monthly Report 1, we will discuss progress and changes in each particular subject area. Much of the content will be narrative, based on our continuing interviews and discussions. Other material will be statistical, based upon data gathered by the local project teams. In noting the consensus of participants we will include discussion of the factors related to progress and the reasons for change. However, the role of Report 1 is historical

rather than analytical, and serious attempts to interpret and predict are reserved for Reports 2 and 3--the analytical portion of our study.

Machines

Ideally, the mechanical characteristics, problems, and complexity of an educational delivery system should be as nearly "invisible" as possible. We are accustomed to broadcast standards that approximate perfect reliability, automatic picture quality, and "idiot proof" operation. This is a difficult target, especially in the light of the elaborate transmission problems that must be overcome.

Although the awesome technology base will not be apparent to the audiences . . .

As discussed previously, we expect an inevitable "debugging" process in the early stages. It is interesting to note that NASA has allowed itself a checkout period for the satellite, but has not scheduled a similar checkout period on the ground. It will be important to know how closely real events follow the initial debugging model we have proposed . . . and whether such problems, if they do occur, are responsible for any initial disillusionment.

Minor difficulties in the man-machine relationship may have profound influence on their attitudes.

Equally important, an educational planner will need to know which problems may be overcome by future developments or changes in operations. For this reason, it is necessary to assess mechanical performance first from the standpoint of the user who is concerned primarily with convenience and picture quality. Secondly, it is necessary to know where problems occur and their general nature, so as to assess the contribution of future developments and the potential for correction.

To do this we will ask each on-site staff to provide a regular report on the picture quality delivered. With the cooperation of the regional activities, we will supply a simple comparison test to indicate picture quality in the classroom. At the same time, we will ask for a minimal evaluation of sound and color quality. Such a simple report is, of course, only one indication of the total system performance, but it may constitute a major issue if any questions of user acceptance arise.

We will gather simple evidence of total system performance

We will also ask for periodic summaries of all equipment malfunctions and servicing required so as to record the sources of any delivery problems as well as their occurrences and the steps taken to correct them. The statistics and discussion of such problems will be noted and described under four categories: (1) the equipment used in generating the original signal, (2) the satellite delivery equipment, (3) the display equipment at the site, and (4) the interactive equipment.

. . . of equipment problems

. . . and their causes,

. . . and what had to be done to restore service.

For purposes of the educational system planner, these details have little significance beyond their effect on overall system performance and user attitudes. Hence, the detail which we will ask is limited, and the frequency of observation will be reduced if it is apparent that performance is adequate to supply a consistent delivery and quality.

Materials

Perhaps the most difficult single aspect of this project will be to separate the character and content of the educational materials from the characteristics of the delivery system. This separation can only be

Our primary concern is the educational delivery system, not the materials.

accomplished by examining the nature of the broadcasts and the way in which they are received. But it must be remembered that the achievement of educational impact is not the goal of the ESCD, and its assessment is ancillary to describing the characteristics of the delivery system. There have been innumerable tests and demonstrations of educational television, and one must not adopt a misguided view that this is another. The results of "asking the wrong question" can only be failure to recognize and assess the true satellite potential.

We are not trying to determine the effectiveness of ETV,

Again, our description will be a mixture of narrative and statistical information designed to provide the future planner and system designer with useful information. Month by month we will note trends and changes in the communication materials, examining them both for their inherent quality and for response to feedback from the target population.

But quality of programs and programming is significant.

For purposes of reporting and measurement we have categorized the materials to be carried on the satellite in four functional areas: (1) education and training; (2) material distribution, such as in the Rocky Mountain area; (3) public information, such as the *Alaska Native Magazine*; and (4) teleconferencing, such as the Appalachian exercises.

Money

Throughout the course of the experiment we will monitor the sequence of expenditure, separating it into the conventional categories of Capital, Operating Costs, and Supporting Costs (consisting of

Cost, of course, is a critical issue.

donated services, costs incurred by affiliated organizations, and intangible costs). The last are, of course, difficult to assess, but nonetheless real in terms of the needs of future planners. We will also attempt to delineate the sources of funds in terms of their origin--in foundation, federal, state, community, or private sectors. We hypothesize that the centroid of funding in a truly viable activity will gradually shift toward the private end of this "spectrum". This tendency probably cannot be verified in a limited demonstration; however, it will be important to document signs of private interest. Already RCA is considering an educational service to Alaska, through its own satellite, with low cost ground terminals similar to those of the ESCD.

Any system must have funds for development, operation, maintenance, and replacement . . . How much is needed, and where can it be obtained?

We have great doubt that this shift can be demonstrated within the time period of this experiment. In the meantime, however, we believe it will be possible to separate costs, both by source and function, so that one may reach some gross cost evaluations for comparison with alternative systems. This comparison process will be under way throughout the project, and because of its importance will form the subject of a separate final report.

What about the costs of alternative delivery systems?

* * * * * * * *

Thus we will continuously monitor the project in terms of these four key factors: Men, Machines, Materials, and Money. As the project progresses we will attempt to structure the information in these four categories as time-lines, partially for our own convenience and partially in the hope that such a display will prove

useful in identifying the three phases, Startup, Adaptation, and Operation, which we have hypothesized.

Clearly Report 1 is essentially historical in nature but designed for a specific application by a future educational policy maker, planner, or system designer. It will differ substantially from reports generated by regional personnel during the pre-launch and project periods. Report 1 will treat each region as uniformly as possible so as to emphasize the potential for comparison and generalization. In this sense Report 1 is designed to present a higher level of summarization and abstraction than the reports generated by the regions. Rather, its preoccupation will be aimed toward issues of inherent interest to the planners of future large-scale educational delivery systems.

Report 2 - MONITORING AND PREDICTING OUTCOMES OF EXPLORATORY SOCIAL PROCESSES: A Case Study of the Operational Phase of the ATS-F Education Satellite Communications Demonstration, 1974-1975

Like a laboratory notebook, Report 2 will document our month-by-month interpretations and predictions.

This report is analogous to a laboratory notebook. In it we will, month by month, document our interpretation of events, our predictions for continued development, and our success or failure in this prognosis. To the extent that Report 2 exhibits valid prognoses, we will have demonstrated a significant insight into the project and its dynamics. Like weather forecasting, the time span of valid forecasts will also reflect the depth of understanding. It is not difficult to predict the weather ten minutes in advance.

Since this process is itself an attempt to interpret and learn from the satellite exploration, we

must expect the initial comments to be tentative and exploratory, following a learning curve toward some useful insight. While Report 2 will be inherently confidential during the course of the program, its content and details will be included in the summary project report so as to provide a basis for independent review and evaluation. Entries in the Report 2 file will be made as often as appropriate, but no less than at monthly intervals. Entries in the log may represent consensus, or they may document the hypotheses and predictions of individual team members. In any case, such generalizations will always be accompanied by predictions which can be verified or negated by actual events.

Report 2 will be confidential during the demonstration.

The method for certifying and protecting the integrity of such records is well established at Battelle. It is detailed in our internal procedures as well as within the laboratory record books themselves. The procedures are particularly stressed because of the periodic need to authenticate such records in patent litigation.

In brief, they require registered notebooks assigned to individual staff members who are responsible for their contents. They must be kept in ink, dated, and signed both by the author and a reliable, as well as knowledgeable, coworker at the time of entry. No pages or parts are allowed to be removed, no erasures may be made, and no blank pages or spaces are allowed. Battelle's laboratory notebooks reflect directly on the integrity of the Institute and are treated with corresponding respect.

Procedures for guaranteeing its integrity are well established.

As indicated in our initial discussion of Battelle's analytical approach, this innovative technique has been developed to cope with the inherently ambiguous character of large-scale experiments like the satellite demonstration. We believe that this technique concentrates on the essential lessons to be learned from such an experiment, and brings to bear the only appropriate and defensible application of scientific method, namely, the ability to make valid predictions. Predictions will be made for all of the categories of input variables. In the category of Men, for example, attempts will be made to predict levels of personnel, and attitudes of implementers and planners of the experiment, as well as factors such as student attendance. In the category of Materials, predictions will be made of changes in program content or structure resulting from feedback of students, teachers, school administrators, etc. Changes in equipment operation and maintenance procedures will be predicted for the Machine category, and changes in the allocations of funds will be predicted for the Money category.

This innovative approach greatly broadens the application of scientific method . . .

To avoid intrusion and experimental "contamination", these predictions will not be shared with participants in the demonstrations until after the operational phase is complete. Instead, Report 2, together with its predictions and the bases for making the predictions, will be recorded in Battelle's standard registered laboratory notebooks. At the end of the project an overall analysis of predictive capability will be made by Battelle and incorporated in the final report together with a copy of the Report 2 documents.

But the predictive process must be isolated from participants in the experiments.

Thus, the "history" documented in Report 2 will consist of the predicted results and the actual

results, along with documented findings intended to account for significant discrepancies between the predicted and actual results. Since Report 2 will be an interpretive effort, only those current or past historical events will be included in this report that are to "explain" discrepancies between predicted and observed results. It is assumed that each region will provide its own detailed documentation of events, so that it is neither necessary, nor desirable, for Battelle to duplicate this effort.

Significant predictive discrepancy will trigger inquiry and improved understanding.

Battelle will use Report 2 predictions in the following manner. Whenever a significant discrepancy is found between a predicted event and the actual event, an "inquiry" will be triggered to determine the possible cause. The inquiries will be carried out through an appropriate combination of the following:

(1) Phone calls made to regional personnel at the level at which the actual event was reported

(2) Regular monthly visits to regional centers

(3) "Random" visits to regional centers.

In no case will phone calls or visits be made to individuals of the target populations. The reasons for this self-imposed constraint are detailed in Appendix D. The purpose of the systematic visits to regional centers is both to obtain material for Report 1 and to obtain first-hand impressions of the actual "environment" within which the reported results are being generated so that correct interpretations of the reported results can be better ascertained. Without such visits it will be impossible to obtain correct impressions of the

We will collect information through visits as well as phone calls and formal data transfer

operational crises, pressures, problems, and general "tone" of the large-scale, multidisciplinary efforts represented by these demonstrations. Such visits will also be necessary to identify similarities and differences among the operational tones of the three regions.

The "random" visits will provide additional assurance of correctly identifying the environmental tone, and will also provide a means for immediate and more detailed inquiry when major discrepancies are identified. We expect that the causes of relatively minor discrepancies will be identified by means of phone calls.

It is emphasized that Battelle will make these regular predictions and compare them with actual results in the most unobtrusive manner possible. Participants will not be given any of the predictive information, and in general, will not know and need not be concerned with this activity.

Battelle's presence will be as unobtrusive as possible.

In Battelle's view this predictive system is a unique and innovative approach to analysis of large-scale social exploration. We believe that it is a significantly improved methodology which, if successfully demonstrated in the ESCD, will constitute a valuable and much needed tool. The tool is inherently designed to test the validity of our insight into social processes and will have application to many other areas of educational policy beside the satellite explorations.

This "predictive analysis" will contribute a significantly improved methodology to large-scale social experiments.

The validity of the procedure rests squarely on a demonstrated ability (or lack of it) in repeated periodic attempts to make valid predictions during an

on-going exploratory social process. We believe that no amount of theoretical reflection concerning the merits of such an innovative monitoring process can substitute for an actual real-world trial. The coming exploration provides an ideal setting for such a trial. Further discussion of the approach is given in Appendix A.

Report 3 - ALTERNATIVE LARGE-SCALE DELIVERY SYSTEMS FOR EDUCATIONAL MATERIALS: Current Experience and Projections through 1985

Ultimately, the educational policy-maker requires a comparison with alternative delivery systems.

Ultimately the interests of a potential policy maker or planner require a valid comparison of the satellite with alternative delivery systems. The exploration and definition of such alternatives, together with collection of the necessary information for comparison, will take place throughout the course of this project. A key factor in such analysis must be the tradeoffs of cost and performance, as well as the philosophic issues associated with the incremental costs of reaching isolated students balanced against the obligations of the educational system in isolated and underdeveloped areas. By developing this definition and analysis of alternatives concurrently with the other parts of the project, we can bring still further insight into the questions which must be answered by the satellite exploration.

We will especially consider hybrid alternatives such as combinations of satellite and cable.

Report 3 will focus on the current and expected future requirements for alternative large-scale educational delivery systems with special emphasis on hybrid mixtures of such systems, such as combinations of cable and satellite systems. The technological

penetrations of many such systems are already publicly visible, and the projection of identifiable trends will be the primary objective of this report.

At the conclusion of the study these delivery and cost alternatives will be summarized as Report 3. As in other areas, the intent will not be one of assessment, but of policy oriented description so that the future planner can weigh the pros and cons of a particular system. This approach is not only appropriate to this particular project; it is a fundamental role of NIE to explore new and alternative methodologies for consideration by local and regional educational planners.

Our approach reflects a fundamental role of NIE to explore alternative methodology for consideration by educational planners.

Summary Final Report

Each of the three reports discussed above will be prepared individually and delivered to NIE as separate volumes. In addition, we will prepare a management-oriented summary which combines the salient information from all three in the most concise and lucid manner. As a synthesis of the three study documents it will fill a particularly useful function--a brief and readable document, suitable for general publication, summarizing the overall experience of the Education Satellite Communications Demonstration.

A management-oriented summary will combine the salient information from our study.

Supplementary Publications

It is Battelle's view that one of the major deficiencies of most large-scale social experiments is the fact that the real lessons to be learned are never published. This stems largely from the fact that neither time nor funding is provided to those individuals who were on the "firing-line" in the experiments.

Supplementary publications will document the experiences and opinions of key participants.

As planners and implementers, such individuals are usually rewarded by what they can accomplish in terms of organizing and directing social processes. They are not typically rewarded by their written contributions to the open literature. By providing such individuals with drafts of concept papers that they can react to, amplify, dispute, or actively collaborate with, and by providing such individuals with some assistance, time, and money, it is Battelle's conviction that extremely valuable facts, beliefs, and convictions will be retained and recorded.

Valuable interpretation is often lost because participants are preoccupied and do not publish.

Such output is typically lost. The "data analyses" that usually result from large-scale social experiments are poor, possibly misleading, substitutions for analyses of the social processes. Over-emphasis on prepost testing, control groups, and experimental designs can completely miss the point of whether or not the social process under exploration uplifts and encourages human aspirations and dignity. Battelle believes that a large-scale social exploration should include invitations to those involved at all levels to comment on related issues of public policy that they or others have identified.

Consequently, we propose to prepare a number of supplementary "concept papers" with the collaboration of individuals associated with the ATS-F demonstration, including site coordinators, state and regional personnel, and individuals from the Federal Government, such as in NIE, OE, and NASA.

Battelle will draft concept papers during the project to elicit and document such insight.

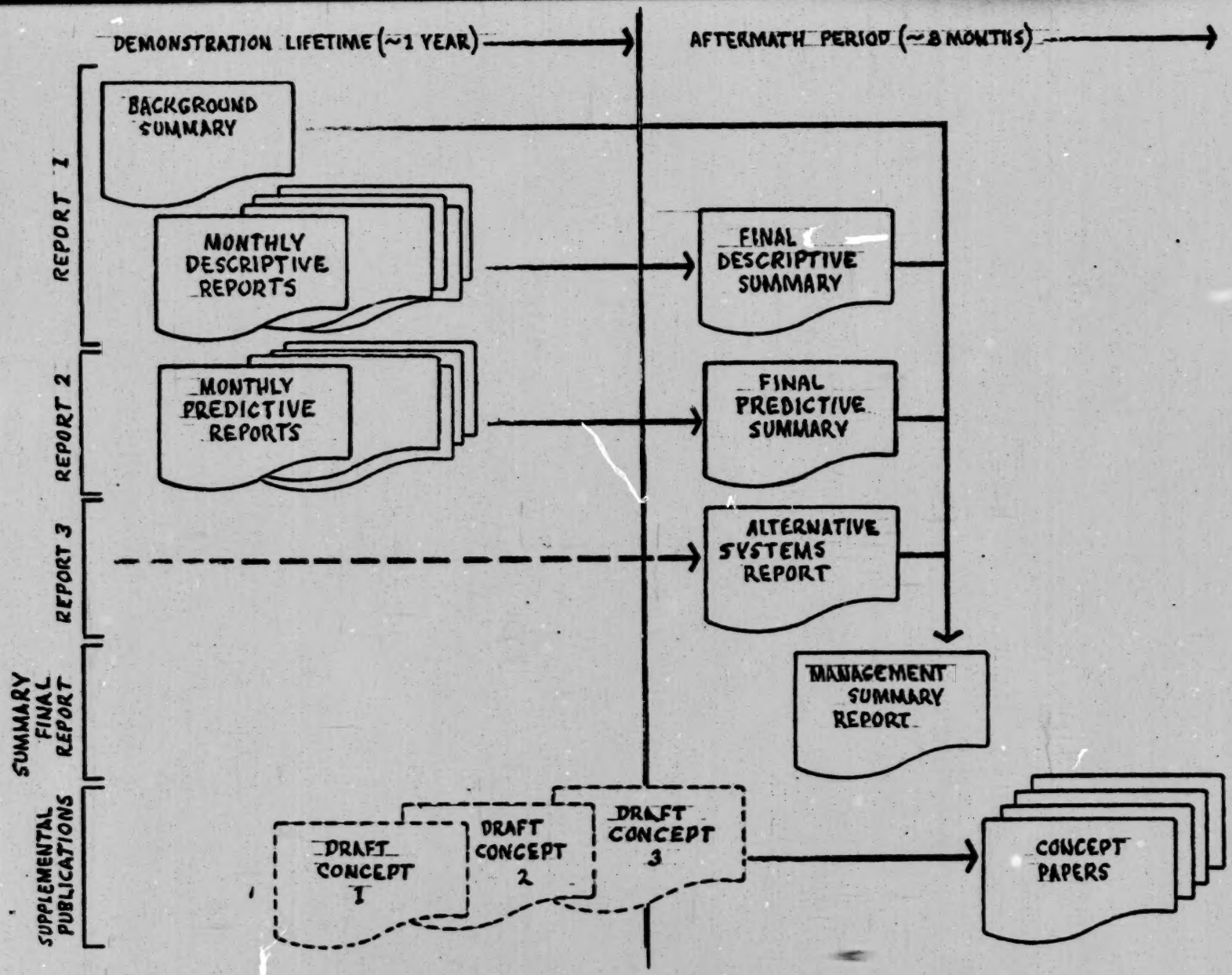
In general, Battelle personnel will draft the concept papers during the operational phase of the ATS-F demonstration, but will invite review, criticism, and/or

collaboration during a 6-month post-operational period. As in the case of information-gathering activities, we propose that support for these contributions be reimbursed through an appropriate addition to the regional funding.

Subject to approval by NIE, these concept papers will be submitted for publication in the open literature. Such publications will assure that the opinions and experiences gained by participants during this program will be available as public information.

These concept papers will be particularly relevant to policy issues.

Tentative examples of the kinds of concept papers proposed for such collaboration are given in Appendix C. The topics proposed for the concept papers focus on a variety of issues concerning public policy and methodology, such as the proper role of the Federal Government in large-scale social explorations, the protection of rights of target populations, and the role of technology in education. If Battelle's proposal is accepted, this listing would be revised as deemed desirable by representatives of NIE and Battelle.



SUMMARY OF PROJECT DOCUMENTATION

REGION: Rocky Mountains

PERIOD ENDING: 30 Sept 1974

Site	Variable being Analyzed[a] Predicted (Basis)		Actual[b]	Difference	Inquiry (Finding)
1	90%	(2)	70%	+20%	Yes (52)
2	80%	(1)	80%	0	No
...					
32	100%	(1)	90%	+10%	Yes (53)
Average over sites	84%	(3)	83%	+ 1%	
Std. dev. over sites	5%	(1)	15%	-10%	Yes (54)

(a) Variable definition: ...

(b) Source: Federation of Rocky Mountain States Document ....

Basis for Prediction:

(1) Same as previous time period

(2) Linear extrapolation over most recent 2 weeks

(3) Quadratic extrapolation

Findings:

(52) Picture quality deteriorated markedly, weather conditions severe.

(53) External cause not determined.

(54) Cultural "holiday" at sites 12, 14, 19, 21, and 23 caused unexpectedly large standard deviation.

A TENTATIVE FORMAT
FOR THE PREDICTIVE ANALYSIS REPORT

ORGANIZATIONAL APPROACH

Project Management Plan

Although our role in the ESCD is clearly that of a nonparticipant, it is also clear that we are not merely reporters. The results of this study will be considered long after the project is complete. Our role is to achieve the best possible understanding of what happened and why, and to present it and other significant factors in a context that can be used by educational systems planners.

In our technical discussion we have described the application of scientific method (our *Mister PV*) to the analysis of a large-scale social experiment (the ESCD) to provide an information base that will be useful to the future planner of educational systems. A formidable array of problems and disciplines are significant to the study.

The scope of our strategy is too broad . . . its role too critical to entrust to a minimal staff, no matter how qualified or experienced. It would be naive indeed to expect two or three people to combine insight, imagination, and experience on such a broad scale. At the same time, because of the ESCD's wide geographic dispersion, only a small group can economically acquire the necessary detailed familiarity with the experimental areas and their activities.

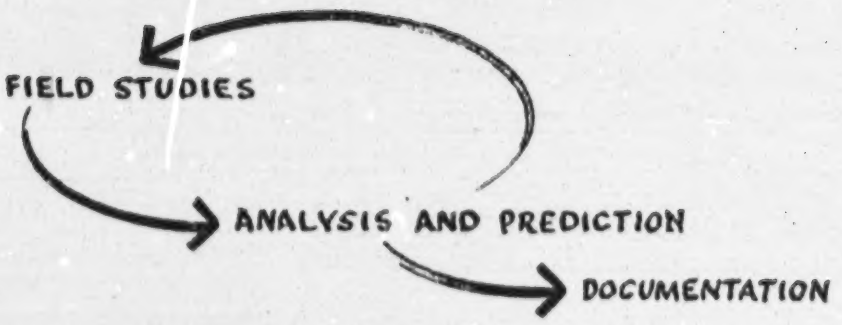
This problem is well recognized at Battelle, and we have developed our management techniques accordingly. We are regularly faced with similar situations in questions of economic assessment, environmental impact, venture analysis, and technology forecasting. We approach these through a combination of concentrated effort by a core group and peripheral involvement by a large contingent representing both diverse specialties and our finest creativity. The ability to provide such support, insight, and ingenuity is one of the key virtues and strengths of Battelle's size and intellectual environment. Phase III activities will revolve about a tightly knit team.

The Project Director, George Tressel, is Manager of BCL's Communication Systems Research Section. He will be especially concerned that the information gathered provides an integrated picture of the three explorations and that the analysis results in a truly substantive basis for future consideration of alternative educational delivery systems. In addition to serving as Project Director, Mr. Tressel will also perform a substantial portion of the research.

A television pioneer, he has been intimately involved in video programming, hardware development, and educational applications since its inception. For his outstanding educational film productions he has twice received citations from the American Society for Information Science, three awards from the Edinburgh Film Festival, and first prize from the Brussels Scientific Film Festival. He led both the recent educational telecommunication policy study and the information system study of integrated television transmission, computer searching, and automated microfiche retrieval. He has described relevant research in such diverse publications as the *Journal of Chemical Education*, the *Journal of the Society of Motion Picture and Television Engineers*, and the *Journal of the American Society for Information Science*.

Assisting Mr. Tressel in management of the project will be the Deputy Director, Phillip Fisher. He will be responsible for the administration and coordination of the program, assuring that the activities are completed within budget and on schedule. Mr. Fisher is Associate Manager of Mr. Tressel's Communication Systems Research Section, where he is responsible for administrative management of all large-scale research programs.

For technical management purposes we have divided the project effort into three primary areas: (1) Field Studies, (2) Analysis and Prediction, and (3) Documentation, interrelated as follows.



(Arrows represent primary information flow.)

As indicated in Figure 3, the work in these areas will be led by George Tressel, Ralph Thomas, and Patricia Brown--the senior staff who prepared the design plan.

Ralph Thomas is Manager of BCL's Statistical and Mathematical Modeling Section. He has authored more than 20 publications on a wide variety of mathematical topics and is especially interested in policy issues related to education, psychology, and management. This interest has led to participation in a number of large-scale social experiments where very large quantities of data and elaborately related variables must be analyzed. Because of his innovative approach to modeling and analysis, he was recently awarded a citation by NASA. His technique was commended for its unusual blend of empirical methods, statistical methods, and analysis of physical constraints.

Patricia Brown is Senior Researcher in the Communication Research Laboratory. She has worked closely with Mr. Tressel in both management and technical investigation of a variety of projects. She is a recognized expert in information systems analysis, design, operation, and documentation, with more than 20 years' experience in systems and survey design, project management, and educational materials development.

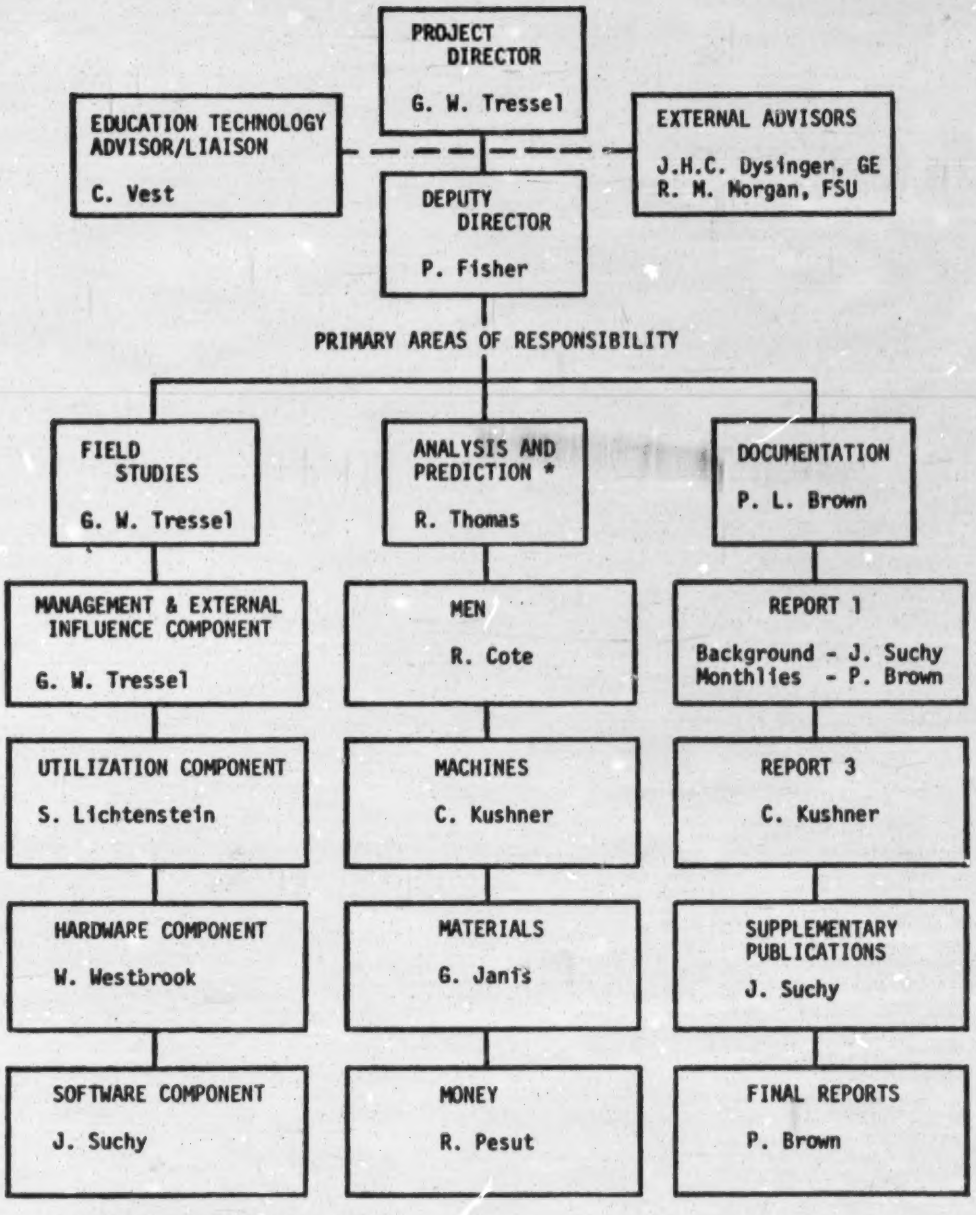


FIGURE 3. PROJECT ORGANIZATION

Miss Brown recently designed the U.S. Army's audio-visual management reporting system and was personally responsible for the design and implementation of a system to assess the effectiveness of the Air Force activities in Southeast Asia. She wrote the National Science Foundation's position paper on information systems for the Stockholm Human Environment Conference and for three years served as Chairman of the National Security Industrial Association's (NSIA) task force on user studies.

Supporting these three principal investigators will be a core group drawn from BCL and Development Associates, Inc. (the same subcontractor firm associated with us in Phase II). Because the three areas are so intimately related, the principal investigators and the core group will be working <u>together</u> on all of them. To ensure that major factors in each are adequately considered, however, we have assigned specific surveillance responsibilities to the core group, as shown in Figure 3. These are <u>responsibilities</u>, not restrictions to perform only in these areas . . . or to be the only performer. This core group will be assisted as necessary by other staff members from Battelle-Columbus and Development Associates. Biographical sketches of team members and representative potential assistants are given in a separate volume, Part B: Statement of Qualifications.

Throughout the program, Mr. Tressel and the project team will work closely with Dr. Vest, whose advice and counsel have proved invaluable in planning this study as well as on previous projects such as the telecommunication policy study. Dr. Vest's prior involvement with the Brazilian satellite experiment and his stature as one of the country's foremost educational technology experts are ideal background for his role as advisor and Washington-based liaison with NIE. He is the Vice Chairman of the NSIA Committee on Educational Technology and served on the planning committee of the first National Conference on the Application of Computers to Training as well as chairing several subsequent conferences on computer-aided training and educational technology.

To add a broader perspective and avoid the pitfalls of "closed group" discussions, we will call upon the advice and counsel of eminently qualified consultants in the eductional technology and satellite application areas. Their long experience in the field will be most helpful in anticipating future trends and key issues in comparing alternative systems.

Work Plan and Schedule

For technical management purposes we have divided the project effort into three primary task areas, scheduled as shown in Figure 4.

- Task 1. Field Studies
- Task 2. Analysis and Prediction
- Task 3. Documentation

Task 1. Field Studies

As we indicated in our proposal for Phase II, and as we reiterate now after our whirlwind visits to the three regions, no amount of documentation review or briefing can substitute for familiarity with the ESCD regional project personnel and detailed first-hand understanding of their plans and problems. Such insight can only develop through visits and dialogue with as much of the cast of characters as possible.

Our predictive analysis approach necessitates monthly visits to each region during the demonstration period. We are well aware that such "visitations" can place an unconscionable drain on the working hours of the place visited, and consequently we have structured our approach around certain basic ground rules that will minimize such problems. Through this approach we expect to develop and maintain a friendly rapport . . . a periodic visitor whose presence is welcomed rather than suffered.

Team Size - No more than a 4-man team.

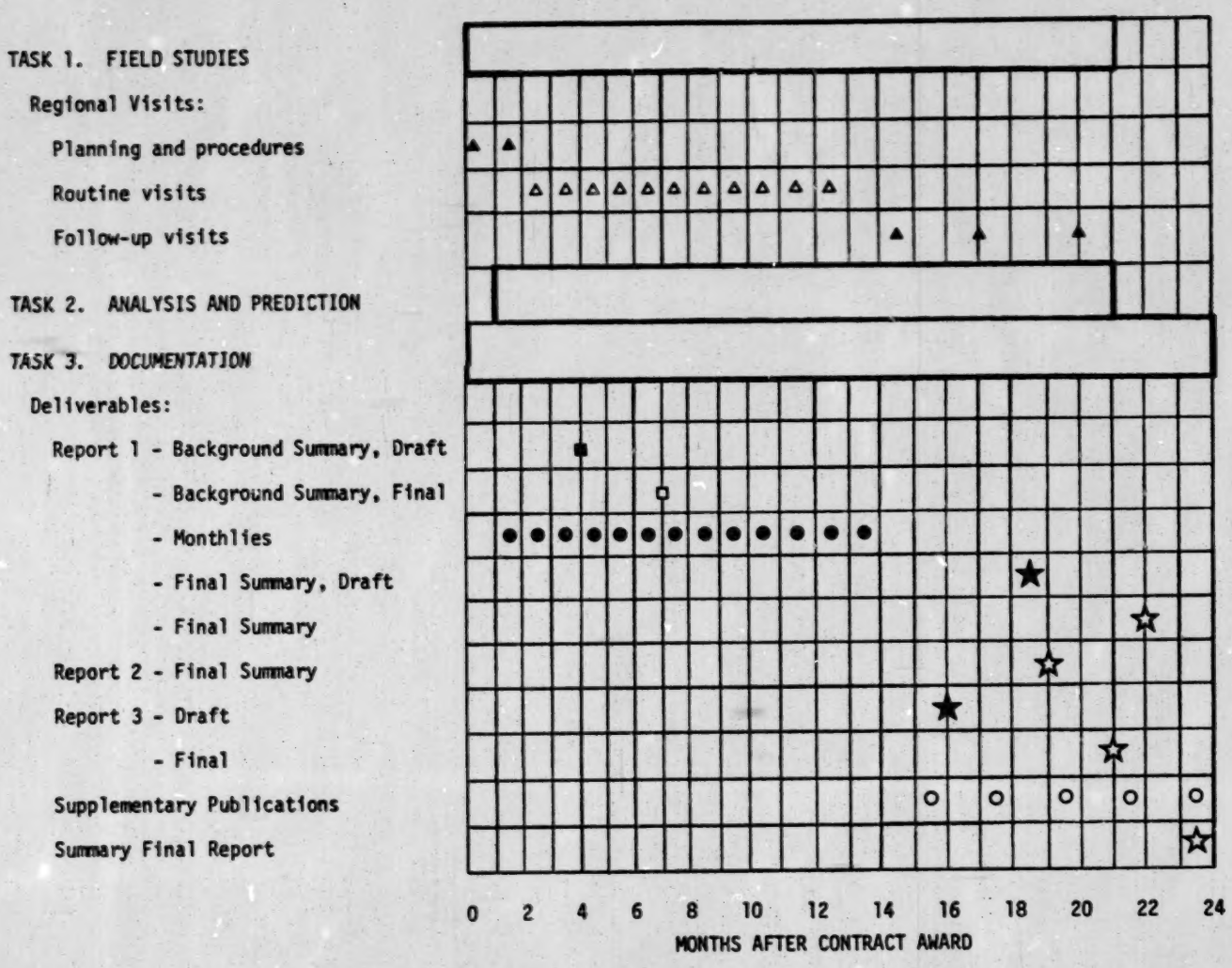


FIGURE 4. PROPOSED PROJECT SCHEDULE

Team Composition - A team leader, representing the Management and External Influence Component. (George Tressel and Ralph Thomas will alternate in this role.)

An information-gathering support staff, representing the other three components--Utilization, Hardware, and Software. (Whether or not three different people visit will depend on the volume and nature of the information needed.)

Team Contacts - Each team member will have previously arranged appointments with those individuals on the project whom they need to see. Team members will not be underfoot except at appointed times.

Whenever possible, we will arrange an after-hours meal conference to meet collectively with the regional staff.

The team leader may also be meeting with appropriate nonparticipants in the community.

Other procedures to insure maintaining our role of observers will be developed in cooperation with each of the regional staffs. We do not believe this will be difficult. In correspondence and conversations we have informed the regional managers of both our general philosophy and the details of our plan as presented in the previous technical discussion. In every case the response has been most encouraging, and there is no question of our ability to carry out this role. Our friendly interest in their problems and progress is already established, and we enjoy a mutual respect and confidence that can only develop between peers with common interests and problems.

The approach to information collection that will be used in the field studies is described in Appendix B. For the first few visits, the primary role of this team is to establish the mechanisms for timely data/

information transfer from the regional teams and from other local sources. After this initial start-up period the field study group will assist as appropriate in this transfer and interview participants and nonparticipants as required. In addition to the monthly trips during the demonstration period, the field study group will make at least three trips to the regions after the demonstration.

Task 2. Analysis and Prediction

Because this task is the core of our plan, it has already been described substantially elsewhere in this proposal. The monthly confidential predictions (Report 2) are a responsibility of this task.

A substantial portion of this activity will involve computer manipulation of data and detailed analyses of the results. We anticipate that during the demonstration period this task will require not only in-depth analyses by key staff but also a number of all-day sessions each month with the entire core group to formulate predictions, discuss discrepancies, and plan inquiries.

Task 3. Documentation

This task concentrates on the products to be delivered to NIE, but it will also be concerned with formatting and systematizing, insofar as practicable, the information recording and flow from Tasks 1 and 2.

The technical report deliverables on this project have already been defined. A tentative submission schedule is shown in Figure 4. The major descriptive reports:

Report 1 - Background Summary

Report 2 - Final Descriptive Summary

Report 3 - Alternative Systems Report

will each be provided in draft form approximately 3 months before the final

version is scheduled, allowing 2 months for NIE and regional review and feedback, and 1 month for subsequent revision.

The schedule of Supplementary Publications (Concept Papers) is indicative only, since the actual number to be prepared and their nature will not be confirmed until after contract award. As stated in the technical discussion, we expect to rough draft these during the demonstration period, then following the demonstration, collaborate with appropriate ESCD participants to revise and submit for open literature publication (subject to NIE release).

In addition to the technical documentation described, each month the NIE will receive from BCL's Project Accounting Office a fiscal report (Statement of Cost) providing a breakdown of actual project expenditures during the previous month.

APPENDIX A

DESIGN AND EVALUATION OF LARGE-SCALE SOCIAL EXPERIMENTS

APPENDIX A

DESIGN AND EVALUATION OF LARGE-SCALE SOCIAL EXPERIMENTS

In virtually all past and current social large-scale experiments, the data obtained do not permit objective analyses leading to valid inferences concerning the initially stated objectives of the experiment. This difficulty is frequently attributed to poor experimental design (lack of control groups, lack of randomization, etc.), poor implementation (lack of suitable training, deteriorations during the experiment, etc.), and poor analyses (especially with respect to the models used, regression artifacts, Hawthorne effects, etc.).

In our view, most of these difficulties are the result of misguided attempts to apply the "scientific method" of the hard sciences to social "experiments". Where homogeneous "materials" can be identified and carefully structured into treatment and control groups, and isolated from extraneous influences long enough, then the word "experiment" has its classical meaning. People are not representable as homogeneous groups, because "matching" can never be achieved on all important attributes.

Not only that, but the basic statistical concept of "population" appears to be threatened, and not applicable to the analysis of social "experiments". For it is a fact of mathematical statistics that for a normal population, an increase in the sample size, n, will result in a decrease in the standard error of the mean by a factor $1/\sqrt{n}$. This useful theory is simply not applicable to large-scale social experimentation, where increases in sample sizes merely serve to introduce new "populations". Said another way, large-scale social experiments really involve "collections of populations", and increasing the size of the sample only increases the size of the collection. Instead of increased precision, one simply increases heterogeneity, and decreases the possibility of making any "overall" conclusions, or generalizations.

We are thus confronted with a virtual "breakdown" of conventional scientific method and conventional statistical methods applied to large-scale social experiments. At the same time there is increasing evidence that our society is becoming an "experimenting society", eager to harness technology to social goals. Certainly such experimentation represents needed explorations of possible solutions to social problems. Such exploration is very costly, however, and unfortunately all too often generates misleading or incorrect information, especially when well-intentioned, but inappropriate, attempts are made to "analyze" social explorations using concepts borrowed from the hard sciences.

In view of these well-documented difficulties, Battelle has developed and applied more fundamental analysis procedures that are capable of yielding scientifically valid results:

Measure, Relate, Predict, and Verify (*Mister PV*)

The "core" of scientific validity involves four processes. In general, quantitative measurements must be made on several different variables; these measurements must then be quantitatively related by graphical, statistical, or mathematical methods. Then, based on previously obtained experience and results, predictions must be made for the results expected at future times or under conditions different from those in current use. Finally, verifications of the predictions constitute the primary evidence of scientific validity.

Battelle calls this procedure the "measure, relate, predict, and verify" approach, or the *Mister PV* approach to scientific validity. The *Mister PV* approach takes the emphasis off of tangential design concepts, such as control groups, randomization, and normality, and instead places emphasis on prediction and verification. The simplicity of this rudimentary approach makes it possible to obtain scientific validity in complex and so-called "unstructured" social experiments. Whenever repeated predictions are consistently and routinely verified, within specified tolerances, then scientific validity has been achieved, whether the "treatment group" consists of a single individual or an entire society.

Despite the apparent simplicity of this approach, most large-scale social experimenters do not undertake repeated, explicit, documented predictions of the results they expect to obtain on the following hour, day, week, etc., during the course of an experiment. By failing to require this, conventional experimental designs frequently fail to yield convincing results.

To the practical person, repeated successful predictions constitute the most convincing evidence that "understanding" really exists. Since a government agency or an entrepreneur may consider making a huge expenditure depending on the outcome of a large-scale social experiment, it is remarkable that the "show me" kind of dynamic prediction and verification process required to justify such expenditures is virtually absent from conventionally conducted social experiments.

A General Description of the Design Approach

To apply this dynamic predictive-verification approach to the design of large-scale social experiments, it is necessary that the design include detailed answers to the following questions:

- What variables are to be measured?
- How are these variables to be measured?
- How frequently are the variables to be measured?
- How are the dependent and independent variables to be quantitatively related?
- How will the quantitative relations be used to make predictions during the course of the experiment?
- How will the predictions be recorded to assure that they are, in fact, predictions?
- How frequently will the predictions be made?
- How frequently will the predictions be verified?
- What criteria will be used to decide whether a prediction has been verified?

It is doubtful that good predictions can be made at the outset of the experiment. If the early predictions are poor, the predictive relations

among the variables would then be reexamined and modified to yield improved predictions for the next stage of the experiment. Ideally, continued iteration during the experiment would yield continued upgrading and improvement in the quantitative relations. Ultimately, good predictions would be repeatedly and consistently made.

In this design approach it is clear that the "number of predictions made" is more fundamental to the design of the experiment than the more customary "number of individuals in the group", or sample size. It is also clear that the design is allowed to "evolve" over time. Variables found to be unimportant or too "noisy" for predictive purposes may be dropped from further measurement. Even the quantitative methods used for generating predictions may change as more data and more experience are gained. In fact, in Battelle's experience, the quantitative relations would be expected to begin with very primitive graphically based predictions with essentially no statistical or conceptual validity. This is the Empirical Stage of the evaluation. However, as more data are gathered and analyzed, and as more predictions are made, it would be expected that certain statistical regularities in the predictive-verification trials would become apparent. At this stage the use of regression or other statistical models is appropriate. This stage of evaluation is the Statistical Stage. If continued gathering and analysis of data suggests relations among "concepts", rather than simple numerical magnitudes, then the final Conceptual Stage of the evaluation is reached.

Thus the quantitative analyses are expected to begin with the Empirical Stage and would, ideally, terminate in the Conceptual Stage where quantitative predictions could be consistently generated and fully explained in cause-and-effect terms, rather than as simple statistical correlations or graphical plots.

It should be noted that this approach does not escape the usual problem of "generalizing" from the existing data to results that might be obtained if the experiment were made with different individuals in a different region, etc. No experimental design escapes this problem. However, the problem of generalization is, itself, a prediction problem. The "learning"

experiences obtained during the successive attempts at prediction would be expected to provide good insights for generalizations. Such generalizations would, of course, be made when possible at the conclusion of the experiment.

APPENDIX B

DATA COLLECTION AND ANALYSIS

APPENDIX B

DATA COLLECTION AND ANALYSIS

We have set interpretive goals that demand an ability to collect, combine, and analyze an unusual range of information. Some of it will be "hard" data, in either raw or partially interpreted form. Other material will be qualitative, a narrative combination of anecdote, assessment, and interpretation. In all cases, we must constantly assess the quality as well as the meaning of the information. Both our collection techniques and our analytical approach have been chosen to provide a means of weighing this quality while allowing flexibility of input. As a result, much of the procedure serves to provide a continual interrogation of the system and a check on the depth and quality of our information. The raw material can be generally described in terms of "numbers" versus "narrative".

Quantifiable Information. This material will be largely obtained from the regional organizations. We propose to compensate them appropriately for the obvious effort, costs, and inconvenience involved in supplying materials as they are gathered. Clearly, for our analytical routines it is necessary to ensure a minimum time lag for data receipt, and we will establish a procedure with the regional groups so that this may be accomplished.

Our project staff will visit the sites at least once each month, and procedures will be geared to this schedule. We will adapt as completely as possible to the formats and routines that the regions have already established, and, in keeping with our philosophy of nonintrusion we will both minimize the amount of effort called for and the extent to which they need be concerned with our activity. The types of quantifiable information to be collected can be categorized as:

- Target population exposure and response
- Equipment performance

- Materials delivery
- Costs
- Staffing

Qualitative and Indicative Information. This type of information will be gathered largely through a program of systematic, minimally structured interviews. Each site will be visited at least once each month to discuss the progress achieved, the experiences gathered, and the participants' interpretation and opinions. These interviews will follow the general content outline of Report 1 but also will serve to probe the reasons for discrepancies between our predictions and actual events. The interviews will be planned so as to provide a range of opinions in the three categories of interest--planners, implementers, and others. At the same time, a variety of related materials such as newspaper editorials, comments in school bulletins, etc., will be collected to supplement and/or corroborate this narrative input.

During most of the project this interview activity will be conducted by various members of our staff, as quietly and unobtrusively as possible. The procedures will be arranged in cooperation with the regional staffs at the beginning of the program so as to assure their understanding of our intent and their cooperation with our approach. At the outset of the project, however, we will conduct all such activities in parallel by more than one staff member ... subsequently checking our individual reports against each other so as to assure that our interview results are compatible and our reporting techniques are consistent.

We have already begun this exercise during the course of the preliminary site visits ... drawing individual observations and assessments; then comparing these individual interpretations. We have found an extremely encouraging consistency, qualified only by the desirability of matching the observer's background with the subject area. Fortunately, we bring to bear both a broad and relevant variety of prior experience and substantial skill. The project staff from both Battelle and Development Associates has been hand-picked for their insight and problem-solving ability, a choice which is reflected in our deviation from perfunctory (and, for this assessment, inapplicable) measurement and analysis.

At the same time, both organizations have current projects in all three areas. As a result, we have excellent and established "listening posts" as well as rapport with the local populations and customs. This empathy and rapport with both the communities and the regional project staffs are critical to our expectation that we can collect such a range of substantive information.

In our visits to the three regional centers we have felt a genuine warmth and sympathetic understanding toward their goals and difficulties. We believe that the feeling has been shared and that we will be recognized as peers and comrades ... who desire neither to glorify nor villify, but rather to record their hard-won progress.

Despite the appropriate academic background and qualifications, our staff is basically pragmatic. We have lived in the classroom, struggled with the problems of educational film and television, and wallowed in the real-world problems of social experiment. We are actively involved in related policy experiments and deeply immersed in similar technical problems. We speak the languages of television, education, psychology, and assessment because these are our problems and careers as well. And for the same reason we will enjoy the confidence of the regional staffs.

The problems of analyzing such varied information and seeking valid interpretations are difficult indeed. Consequently, much assessment of such large-scale social experiments avoids the true issues of policy and instead concentrates a barrage of statistical effort on restricted and questionable targets. Critical insight into the key issues of such ambitious undertakings is usually reserved for the period after the conclusion of the project. At this point, any analysis is retrospective at best ... beyond the pale of further testing or verification and a subject for endless debate. Looking backward, with only limited information, at bland and questionable statistics ... and without the ability to probe further ... such "analysis" can be better described as *rationalizing*. A principal virtue of the predictive analysis system is the fact that it will both test our insight into available data and trigger a variety of salient questions to be explored while the system is still operational. It is because of this evolutionary characteristic that we expect to generate a deeper level of information and understanding

than would be possible by purely retrospective analysis. Each discrepancy between prediction and fact may pose a new question, and the end result of this iterative analysis and information-gathering procedure should be a far more valid and useful picture than would otherwise appear.

Fortunately, we are also among the pioneers who are exploring the application of the AID program to such difficult issues. It is our experience that these techniques can expose provocative relationships between tenuous and diverse input and output variables. But perhaps more important, the program has often proved equally powerful in identifying a lack of relationship or an inadequacy of the measurement technique.

Where assessment and opinion are involved, the initial information is of great question--"big and small", "better or worse", "acceptance and rejection" mean different things to different people. It is, of course, for this reason that we choose to make such judgments with a controlled staff group whose semantics can be standardized. But more important still, techniques like the AID program can provide some concrete indication of relevance to real events.

We do not at this juncture propose the details of our analytical activities. We have chosen the term "exploration" advisedly, and our approach must be sufficiently flexible to follow the real course of events. However, the general results from our analyses may be anticipated somewhat. It may be expected, for example, that an initial debugging period, of possibly a month, will occur at the outset of the operational phase. The data obtained in this period on output variables and attitudes are not likely to be useful for making predictions. However, they may help indicate when the debugging period may be considered as ended and a steady-state period considered to have begun. The best predictions and trends in output measures and attitudes are likely to occur in the steady-state period. The data obtained in this period are expected to be of the most importance. A final "close-out" period, of possibly a month, may cause the data near the termination date to be unreliable because of problems created by closing down the demonstration.

After the termination of the demonstration program, Battelle proposes to make summarizing analyses of all of the data obtained during the

program. Particular emphasis will be placed on all of the data obtained during the steady-state period. Extrapolations of trends and attitudes during this period may yield the best predictions of what would be expected if the program had continued indefinitely.

SUMMARY

Battelle asserts that no one truly understands social processes, especially exploratory social processes. We believe that part of the failure derives from the fact that neither individuals nor institutions typically exhibit the self-discipline required to document and verify their insights. We believe that a useful device for attacking this problem consists of the simple requirement that predictions be made of future outcomes of social processes. If the predictions are correct an understanding is demonstrated; if the predictions are not correct, there is something to be learned. Either way, the "exercise" of prediction is beneficial. If successful, we believe that this "predictive analysis" innovation will establish a new standard for the scientific study of social processes.

APPENDIX C

POTENTIAL CONCEPT PAPERS

APPENDIX C

POTENTIAL CONCEPT PAPERS

1. *Towards a Definition of "Successful" Social Experimentation: The ATS-F Experience*

This concept paper would explore the possibility that a social experiment may be declared "successful" if the planners, implementers, and target groups believe that the experiment should be continued indefinitely, provided the cost is gradually transferred to the target group.

2. *On Defining a Suitable Role for Federal Agencies in the Implementation and Evaluation of Large-Scale Social Experiments: The ATS-F Experience*

This concept paper would explore the idea that a proper role for NIE, for example, would consist of funding exploratory programs involving the widest possible variety of alternative large-scale educational systems. As long as our society consists of widely diverse cultural groups, no single educational system is likely to be "optimal"; yet practical experience must be gained if the potential benefits of differing educational systems are to be conclusively identified.

3. *Are Rights of Target Populations Legally Defined and Protected in Large-Scale Social Experiments?: The ATS-F Experience*

This paper would explore the possibilities that the rights of target populations are <u>not</u> legally defined and protected in large-scale social experiments. In the ATS-F experience unique cultural groups are to be exposed to varying degrees of "technological intrusion". On the one hand it may be argued that, in the long run, every "corner" of our society will be immersed in technology. If it will happen eventually, why not maximize its rate of occurrence? At the opposite extreme, it may be argued that the existing unique isolated cultural groups must be preserved, so that technological intrusion should not be permitted. Carried to its logical extreme, this approach would result in the creation of human zoos protected from technological intrusion for the benefit of future

historians and anthropologists. The "proper" rate of technological intrusion probably lies between these two extremes. If so, how is an appropriate rate to be determined?; what role will the target population have?; how can the rights of the target population be determined and enforced?

4. *A Method for Social Self-Discipline: Can Society Monitor and Learn From Its Own Behavior?*

This paper would be concerned with the monitoring systems and the associated predictive mechanism proposed for trial use by Battelle during the ATS-F demonstration. It is anticipated that the proposed procedure will, in fact, yield the following benefits:

(1) Frequent periodic predictions, with attempted verifications, will approximate an "immediate response" situation, providing a rapid and improved method for the social system to "learn" its own characteristics.

(2) For those behavioral characteristics that are found to be predictable, for practical purposes an "understanding" of the system behavior is thereby demonstrated.

(3) For those characteristics that are found to be unpredictable, two alternatives follow: (a) by detecting the discrepancy early, it is much more likely that the cause of the discrepancy can be identified, and possibly corrected for, in future predictions; (b) if repeated predictive failures occur, it may be concluded that the corresponding characteristic simply cannot be predicted, and the characteristic should be treated as random in future designs and plans.

(4) The underlying principle consists of the possibility that some social characteristics are predictable and some are not. The identifications of which are and which are not can only be accomplished by a documented procedure of making predictions and attempting to verify them at frequent periodic intervals.

(5) The primary difficulty with current practice is seen to be due to the fact that "data are perishable". No one likes to analyze "old" data; it is impossible to find out what "old" numbers or statements really meant at the time they were recorded. In contrast, it is usually possible to find out what recently recorded data mean; and the recorded data, together with their meaning, can be used to make valid extrapolations to a future time period.

(6) If these assertions are true, the lack of understanding of social systems, and the very need for social experimentation, may be largely due to the fact that the social self-discipline required for making documented predictions has not been advocated and given a fair trial.

5. *When Should Federally Funded Social Experiments Be Terminated?*

This paper will explore several concepts. It may be argued that there are several areas in which the Federal government will be funding programs forever. These include, for example, Welfare and Disabled Veterans. It is not clear, however, whether the Federal government should be in the "education business" forever. One possible mechanism for imposing a limitation on the duration of a federally funded social experiment is the following. Before approving a funded program the government could require a prediction on the part of the contractors as to when non-Federal funds would be available for continued support. Frequent, periodic updating of these predictions, together with the supporting basis for the predictions, should be required. The documented inability to make valid predictions of the timing of non-Federal support may then serve as a basis for terminating support.

6. *Is Education, Without an Intense Technological Exposure, an Educational Handicap?*

It is apparent that our society is becoming immersed in technology. The proliferation of communication and computer networks suggests that the question of the "impact of technology on education" is improper. Technology is becoming part of the "environment". A student who does not have "hands-on"

experience with advanced communication and computer equipment may be, to some extent, uneducated. It may be argued that if he is uncomfortable in the technological world, and is not capable of using the technology networks to increase his own capabilities and satisfactions in life, then some kind of failure in his educational background has occurred. This paper will develop this theme.

APPENDIX D

TARGET POPULATIONS AND THE "ASSESSMENT" PROCESS

APPENDIX D

TARGET POPULATIONS AND THE "ASSESSMENT" PROCESS

In this, as in any substantial exploration, the effects of the experiment itself cannot be ignored. The project must inevitably cause a significant change in the cultural and social, as well as educational, milieu, and once mounted, its effect can never truly be withdrawn. This and many related philosophical issues must be addressed in the course of this project.

In the meantime, one must approach such intrusion with great respect and care. The ESCD exploration has already injected a substantial increase in technology as well as instituted changes in the roles of teachers and in the organization of school components. The assessment teams of each region are keenly aware of this intrusion. In each case they have indicated a serious concern for the effect of additional observers, and a hesitancy lest the observers "outnumber" the target population or enter the scene without regard for any effects on the experimental outcome.

We at Battelle thoroughly agree with this concern and are particularly sensitive to the problems that will be engendered if it is ignored or circumvented. We propose that there be essentially no direct contact between our staff and the immediate, regional target populations. The presence of one assessment team is a necessary evil; the presence of a second would be inexcusable. This is more than a question of Hawthorne effect, or even of empathy with the target population (not to mention logistics); it is a question of maintaining the valid and appropriate role of Battelle as a neutral observer.

We have indicated previously that it is our goal to provide a neutral overview of the program, and we have no interest in measuring its "success or failure" (in our opinion these terms are meaningless) but rather in collecting and interpreting the experience for future planners. This role of neutral observer cannot be exercised by the participants who are

deeply concerned with planning, funding, controlling, and implementing the experiment, "tuning" its activities from day to day to achieve maximum impact within the time and resources available.

On the other hand, this is the only "assessment" role which cannot better be fulfilled by the participants themselves. And it is for this reason that we believe contact with the target population is the proper domain of the regional project teams alone. We suggest that the ESCD has already experienced the effect of such confusion of sponsor-implementer-evaluator roles, and the result has been a significant stress within the program.

We are sure that direct contact with the population is not only undesirable, it is also unnecessary. The future planner considering alternative delivery systems will seek a mature evaluation of the experience, not a detailed measurement of the materials transmitted. Some kinds of material must be delivered in this exploration, and their variety of intent and character contribute to our understanding of potential applications. But for this purpose, detailed evaluation of individual programs is, to a large extent, irrelevant, and the highly qualified and professional regional assessment teams are more than competent to provide the answers to such questions as may arise. They are also both dedicated to their task and sensitive to its magnitude, and additional presence in the field would be a totally misguided multiplication.

As the representatives of a future planner we will have far more interest in the secondary impact engendered by the simple presence of such awesome new technology, ... interest in the debating process within the community--the pros and cons and "conventional wisdom" of those at some moderate distance from the classroom.

We are ideally situated to receive and interpret such messages. Both Battelle and Development Associates have current activities in all three regions, and, in this sense, we already have a network that can furnish feedback regarding local attitudes and discussions. The association between Battelle and Development Associates was established especially to

explore this potential, and we believe that it will provide more and better insight than direct and heavy handed attempts at "measurement" by outsiders.

Finally, the presence of "outside" measurement teams could itself stimulate an unfortunate reaction. A tightly knit rural ethnic population, already the subject of probing and exploration, will stand only so much examination and testing before it turns on the outsider. The margin of balance between friendly acceptance of "improved education" and bristling rejection of "outside interference" is not as great as one might wish.

It is only through sensitivity to such problems that we expect to avoid them. We consider that isolation from direct activity with the target population is a minimal and acceptable step in this direction.